Copy No. _____

Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

Adaptive target classification with imaging sonar as a partially observable Markov decision process

Vincent Myers

Defence R&D Canada – Atlantic

Technical Memorandum
DRDC Atlantic TM 2009-154
January 2010

Canada

This page intentionally left blank.

Adaptive target classification with imaging sonar as a partially observable Markov decision process

Vincent Myers
DRDC Atlantic

Defence R&D Canada – Atlantic
Technical Memorandum
DRDC Atlantic TM 2009-154
January 2010

Principal Author

Original signed by Vincent Myers

Vincent Myers[DRDC Atlantic]

Approved by

Original signed by David Hopkin

David Hopkin
Head/MAP Section

Approved for release by

Original signed by Ron Kuwahara for

Calvin Hyatt
Head/Document Review Panel

Abstract

Discriminating between different types of objects on the seabed using high-resolution imaging sonar is challenging, in part due to the inference of a 3-dimensional shape using one or more 2-dimensional projections. Often, more than one aspect of a detected object is required in order to correctly classify it as one of a number of targets of interest such as an influence mine, or as a benign non-target. When the sensor is equipped by an autonomous underwater vehicle (AUV), the acquired aspects are usually determined by a pre-planned mission rather than in response to the sensor data (i.e. the process is purely deliberative rather than reactive). In this paper, the multi-aspect target classification problem is modeled as a partially observable Markov decision process (POMDP). The solution to a POMDP is called a policy which provides a means for an AUV's control system to determine a course of action in response to the incoming sensor data, such as classifying an object as a target or not, or obtaining data from a specific aspect to reduce the uncertainty. The components of a POMDP for multi-aspect object classification with an AUV equipped with a sidescan are formulated. A numerical simulation is undertaken to assess the performance of the resulting policy and the vehicle path which is reacting to simulated sensor data is shown. The simulation was performed using two targets (a cylindrical object and a small, wedge shaped object) and a non-target class modeled as an elliptical object with uniformly distributed major and minor axes. The resulting policy was executed for many iterations and compared to one which obtains two perpendicular aspects. The classification accuracy of the POMDP model, as measured using a confusion matrix, was markedly superior to the cross-hatching strategy, while obtaining an average of 2.22 aspects (versus 2 for cross-hatching).

Résumé

La discrimination des divers types d'objets sur le fond marin au moyen de sonar d'imagerie haute résolution représente un défi, en partie en raison de l'inférence d'une forme tridimensionnelle partir d'au moins une projection bidimensionnelle. Il faut souvent plus d'un aspect d'un objet détecté pour classifier celui ci correctement parmi diverses cibles d'intérêt, comme une mine à influence, ou un objet inoffensif ne constituant pas une cible. Lorsque le capteur est doté d'un véhicule sous marin autonome (AUV), les aspects acquis sont en général déterminés au moyen d'une mission planifiée, plutôt qu'en réponse aux données du capteur (c. à d. le processus est strictement délibératif au lieu de réactif). Dans le présent document, le problème de la classification de cibles à aspects multiples est modélisé comme étant un processus décisionnel de Markov partiellement observable (PDMPO). La résolution d'un PDMPO est appelée une politique, lequel fournit un moyen pour le système de commande de l'AUV de déterminer un plan d'action en réaction aux données provenant des capteurs, notamment la classification d'un objet comme une cible ou un objet ne constituant pas une cible, ou l'obtention de données à partir d'un aspect

en particulier en vue de réduire l'incertitude. Les éléments d'un PDMPO pour la classification d'objets à aspects multiples au moyen d'un AUV muni d'un système à balayage latéral sont présentés. Une simulation numérique a été entreprise pour évaluer la performance de la politique résultante, et le trajet du véhicule, lequel réagit aux données simulées du capteur, est présenté. La simulation a été effectuée au moyen de deux cibles (un objet cylindrique et un petit objet en coin) et d'une classe d'objets ne constituant pas des cibles, modélisées comme un objet ovale comportant des axes principaux et secondaires répartis uniformément. La politique résultante a été réalisée à de nombreuses reprises et comparée à la politique qui dicte l'obtention de deux aspects perpendiculaires. La précision de la classification du modèle PDMPO, comme mesuré au moyen de la matrice de confusion, était nettement supérieure à celle de la stratégie des aspects perpendicularies et permettait d'obtenir une moyenne de 2,22 aspects (par rapport à 2 pour les aspects perpendiculaires).

Executive summary

Adaptive target classification with imaging sonar as a partially observable Markov decision process

Vincent Myers; DRDC Atlantic TM 2009-154; Defence R&D Canada – Atlantic; January 2010.

Background: The problem of multi-aspect classification of targets in high-frequency sidescan sonar images with an autonomous underwater vehicle (AUV) is examined. Discriminating between different types of objects on the seabed using an imaging sonar is challenging, in part due to the inference of a 3-dimensional shape using one or more 2-dimensional projections. Often, more than one aspect of a detected object is required in order to correctly classify it as one of a number of targets of interest such as an influence mine, or as a benign non-target. Typically, the Concept of Employment (CoE) for sidescan sonar missions with autonomous underwater vehicles (AUVs) is to obtain two perpendicular aspects of every part of the seabed (called cross-hatching) to obtain a satisfactory probability of detection. This is sometimes followed by a classification step where a pre-planned multi-aspect survey is carried over individual objects to systematically obtain a number of different aspect angles with little regard to how those are incorporated into the classification strategy, or whether or not they are actually required. This paper addresses the multi-aspect target classification problem using a framework called a partially observable Markov decision process, or a POMDP. Essentially, a POMDP is a general model for an agent which interacts with an environment and tries to maximize some kind of reward (i.e. a *rational* agent). The agent is able to take a number of actions and can be in any number of states, and the reward depends on which action is taken in which state. The process becomes partially observable when the agent cannot directly know the state it is in except through a number of observations which are a function of the unknown state. The solution to a POMDP is called a policy, which is a method of computing the best action to take based on which state the agent believes it is in.

Principal results: The agent, here an AUV equipped with a sidescan sonar, interacts with its environment through a series of observations which are the sonar images, or more specifically the features derived from them. The states are the possible target and non-target classes. The actions that the AUV can take are to classify the object under investigation as one of the known target or non-target classes, or to obtain a new aspect. The AUV obtains a reward for a correct classification and a penalty for an incorrect one, and there is also a cost associated with obtaining additional aspects. The POMDP solver must consider immediate rewards as well as long-term ones. A simple observation model was created based on a single feature, namely the measured length of the object under consideration at the given aspect. A numerical simulation was carried out using two targets (a cylindrical object and

a small, wedge shaped object) and a non-target class modeled as an elliptical object with uniformly distributed major and minor axes. The resulting policy was executed for many iterations and compared to one which obtains two perpendicular aspects. The classification performance of the POMDP model, as measured using a confusion matrix, was markedly superior to the cross-hatching strategy, while obtaining an average of 2.22 aspects (versus a fixed number of 2 for the cross-hatching).

Military significance: In addition to the increased classification performance, the main advantage of the method developed here is that it can supply a vehicle's control system with waypoints to follow for a *reactive* strategy for multi-aspect classification, rather than a purely *deliberative* method based on pre-planned missions. The result is better overall mission performance with a potential for increased operational tempo and can enable an adaptive, reactive capability for unmanned autonomous control systems which have so far been prominently of the deliberative kind.

Future work: This technique must be validated using real sonar data. The sonar data, however, must be of sufficient quality as to be able to properly quantify the variation with aspect of the targets under consideration. To this end, a set of high resolution synthetic aperture sonar (SAS) data is being obtained from the NATO Undersea Research Centre. While much of the POMDP framework developed here will remain the same, a more robust model-based feature is being developed as the observation model. The policies computed by the solver will then be evaluated. If the technique is suitable, a longer term objective is to integrate a version of this technique into the control system of a vehicle such as the Interim Remote Minehunting and Disposal System or another suitable platform equipped with an imaging sensor.

Sommaire

Adaptive target classification with imaging sonar as a partially observable Markov decision process

Vincent Myers ; DRDC Atlantic TM 2009-154 ; R & D pour la défense Canada – Atlantique ; janvier 2010.

Introduction : Le problème de la classification d'aspects multiples de cibles dans les images de sonar à balayage latéral à haute fréquence au moyen d'un véhicule sous marin autonome (AUV) est examiné. La discrimination des divers types d'objets sur le fond marin au moyen de sonar d'imagerie représente un défi, en partie en raison de l'inférence d'une forme tridimensionnelle à partir d'au moins une projection bidimensionnelle. Il faut souvent plus d'un aspect d'un objet détecté pour classifier celui ci correctement parmi diverses cibles d'intérêt, comme une mine à influence, ou un objet inoffensif ne constituant pas une cible. En général, le concept d'emploi (CE) pour les missions réalisées au moyen d'un sonar à balayage latéral comportant des véhicules sous marins autonomes (AUV) consiste à détecter deux aspects perpendiculaires de chaque pièce sur le fond marin en vue d'obtenir une probabilité de détection satisfaisante. Cette étape est parfois suivie d'une étape de classification dans laquelle un levé planifié d'aspects multiples est réalisé sur des objets individuels pour obtenir systématiquement divers angles de cibles sans tenir compte de la faon dont ces angles sont intégrés dans la stratégie de classification, ou s'ils sont vraiment nécessaires. Le présent document porte sur le problème de la classification de cibles à aspects multiples au moyen d'un cadre appelé processus décisionnel de Markov partiellement observable (PDMPO). En somme, un PDMPO est un modèle général pour un agent qui interagit avec un environnement et essaie de maximiser une certaine forme de récompenses (c. à d. un agent rationnel). L'agent peut prendre diverses actions et être dans de nombreux états, et les récompenses varient en fonction des actions prises pour un certain état. Le processus devient partiellement observable lorsque l'agent ne peut connatre directement l'état dans lequel il se trouve, sauf à partir de diverses observations qui dépendent de l'état inconnu. La solution d'un PDMPO est appelée une politique, soit une méthode pour calculer la meilleure action à prendre en fonction de l'état dans lequel l'agent croit se trouver.

Résultats : L'agent, dans le présent cas, un AUV muni d'un sonar à balayage latéral, interagit avec son environnement à partir d'observations qui sont les images sonar, ou plus précisément des caractéristiques obtenues à partir de ces images. Les états sont les classes possibles de cibles et d'objets ne constituant pas des cibles. Les actions qui peuvent être prises par l'AUV consistent à classifier l'objet examiné parmi les classes connues de cibles et d'objets ne constituant pas des cibles, ou à obtenir un nouvel aspect. L'AUV obtient

une récompense pour une bonne classification et une pénalité pour une classification erronée, et un coût est en outre associé à l'obtention d'aspects supplémentaires. Le résolveur PDMPO doit tenir compte des récompenses immédiates ainsi que des récompenses à long terme. Un modèle d'observation simple a été créé à partir d'une seule caractéristique, en l'occurrence la longueur mesurée de l'objet examiné sur un aspect donné. Une simulation numérique a été effectuée au moyen de deux cibles (un objet cylindrique et un petit objet en coin) et d'une classe d'objets ne constituant pas des cibles modélisées comme un objet ovale comportant des axes principaux et secondaires répartis uniformément. La politique résultante a été réalisée à de nombreuses reprises et comparée à une politique pour lequel deux aspects perpendiculaires ont été obtenus. Les performances de classification du modèle PDMPO, comme mesuré au moyen d'une matrice de confusion, étaient nettement supérieures à celles de la stratégie des aspects perpendiculaires et permettaient d'obtenir une moyenne de 2,22 aspects (par rapport à 2 pour les aspects perpendiculaires).

Portée : La méthode développée dans le cadre de la présente étude, en plus d'accroître les performances de classification, a pour principal avantage de fournir un système de commande de véhicules comportant des points de cheminement à suivre pour une stratégie réactive visant la classification d'aspects multiples, contrairement à une méthode strictement délibérative axée sur les missions prévues. La mission offre ainsi de meilleures performances globales permettant un accroissement possible du rythme opérationnel, et elle fournit une capacité d'adaptation et de réaction aux systèmes de commande autonome sans pilote, lesquels, jusqu'à présent, étaient surtout de type délibératif.

Recherches futures : La présente technique doit être validée à partir de données sonar réelles. Ces données doivent, cependant, être d'assez bonne qualité pour quantifier correctement la variation des aspects des cibles examinées. à cette fin, un ensemble de données sonar à ouverture synthétique (SAS) à haute résolution a été obtenu du Centre de recherche sous marine de l'OTAN (NURC). Bien qu'une partie du cadre du PDMPO mis au point dans le contexte de la présente étude reste inchangée, un élément plus solide axé sur le modèle est développé comme modèle d'observation. Les principes calculés par le résolveur seront ensuite évalués. Si la technique est appropriée, un objectif à plus long terme est d'intégrer une version de cette technique au système de commande d'un véhicule comme le Système télécommandé provisoire de chasse aux mines et de déminage ou une autre plate forme adéquate munie d'un capteur d'images.

Table of contents

List of figures

1 Introduction

The automatic detection, classification and identification of objects such as mines on the seabed using high-frequency imaging sonar presents some remarkable challenges, many of which stem from the difficulty of inferring a 3D shape from a 2D projection of an object [1]. This mine coutermeasures (MCM) activity is now increasingly being performed by unmanned systems, such as autonomous underwater vehicles (AUVs), which provide the ability to keep personnel and ships out of areas of danger. More and more, focus is moving away from the purely *deliberative* model for AUVs, where pre-planned missions are prescribed by a human operator, and turning towards incorporating some measure of *reactive* capability, allowing the vehicle to respond to incoming sensor data or to adapt to changing environmental conditions and mission objectives. The objective is better mission performance and increased operational tempo.

The sensor most commonly used for integration with AUVs for MCM operations is the real or synthetic aperture sidescan sonar. This sensor creates a strip-map acoustic image of the seafloor by emitting high-frequency sound in the direction perpendicular to the vehicle track; the resulting image appears like an aerial grayscale photo of the seafloor which is illuminated from the side and an observer's point of view from the top. The achievable resolution of most sidescan sonars, particularly using synthetic aperture technology, is such that is is usually possible to detect targets including modern low target strength mines; however, with this capability comes an increase in the number of naturally occurring clutter objects such as rocks which are of the same general size and shape as the targets of interest. At the frequencies considered here (generally $\geq$ 100 kHz), there is very little penetration of the transmitted sound into the objects which can be used to determine internal structure, and therefore features used for classification are generally computed on the projected geometric shape of the object. Given the higher density of clutter objects now considered, it is increasingly likely that non-targets may appear target-like from any single or combination of aspects. There have been numerous studies on how to combine several aspects of the same target in order to increase the correct classification rate, for instance [2], [3], [4] and [5]. In these studies, the aspects used for classification are fixed and depend on the pre-planned search pattern which was programmed into the vehicle before the survey; typically, this is a back-and-forth "lawn-mower" pattern designed to search an entire area in one search direction, or two perpendicular directions. This is sometimes followed by a "classify" pattern, meant to obtain a large number of aspects (e.g. in increments of 10^{o}) over a single detected object in order to better ascertain its nature and increase the probability of classification. The objective of this paper is to develop a technique not only for classifying targets using multiple aspects but also a way of dynamically determining, based on the sensor data, which aspect the vehicle should obtain next in order to more quickly classify the target under consideration.

The method used here is to consider the multi-aspect mine classification problem as a

partially observable Markov decision processes (POMPD). POMDPs have been applied to several problems in many fields, including robot navigation and target identification [6]. The problem of angle-dependent classification of targets by a robot for mine identification was addressed in [7] for land mines. The model can account for not only uncertainty in the class of the object, but also uncertainty in the features computed from the sensor data. Essentially, a POMDP is a general model for an agent which interacts with an environment and tries to maximize some kind of reward (i.e. a *rational* agent). The agent is able to take a number of actions and can be in any number of states, and the reward depends on which action is taken in which state. The process becomes partially observable when the agent cannot directly know the state it is in except through a number of observations which are a function of the unknown state. The solution to a POMDP is called a policy, which is a method of computing the best action to take based on which state it believes it is in. This policy provides a means for an AUV's control system to determine a course of action in response to the incoming sensor data, such as classifying an object as a target or not, or obtaining data from a specific aspect to reduce uncertainty.

The background of POMDPs is reviewed in Section 2. Then, in Section 3 the multi-aspect object classification problem using a sidescan sonar equipped AUV is posed in terms of the components of this POMDP. A simple observation model was created based on a single feature and a numerical simulation is carried out using two targets (a cylindrical object and a small, wedge shaped object) and a non-target class modeled as an elliptical object with uniformly distributed major and minor axes. The resulting policy is tested over many iterations with random targets and non-targets, and compared to a strategy which obtains two perpendicular aspects using a confusion matrix. Some conclusions based on the simulation are drawn in Section 4.

2 POMDP Background

This section provides some of the mathematical background on POMDPs which is needed to formulate the multi-aspect target classification problem [8]. A Markov decision process(MDP) is a model for an agent which is interacting within an environment and is made up of four components [9]:

1. a finite set of states $\mathcal{S}$,
2. a finite set of actions $\mathcal{A}$,
3. a state transition matrix $T : \mathcal{S} \times \mathcal{A} \mapsto \Pi(\mathcal{S})$ which maps states and actions to probabilities over states. $T(s,a,s')$ is defined as the probability of ending up in state s' given that the current state is s and action a is taken.
4. a reward function $R : \mathcal{S} \times \mathcal{A} \mapsto \mathbb{R}$ which defines rewards for taking each action in each state. The function $R(s,a)$ the reward for taking action a in state s. Note that the reward is only dependent on the current state, which is the *Markov* property of the system.

A solution to an MDP is in the form of a policy π which describes the behaviour of the agent and $\pi(s)$ determines the action to take when in state s. The agent acts in a rational way as to maximize its expected reward over time, balancing long term rewards with immediate gains. The expected future discounted reward over k steps is:

$$E\left[\sum_{t=0}^{t=k-1} \gamma^t R(s_t, a_t)\right], \tag{1}$$

where $0 < \gamma < 1$ is a discount factor which regulates how future rewards are factored in favour of immediate ones. The variables s_t and a_t represent the state and actions at time t and k defines the horizon (the total possible number of time steps) of the MDP. A large discount factor gives future rewards greater effect on the current action.

Given a policy π, the value $V_{\pi,t}(s)$ of state s at time t for is the expected discounted accrued reward defined recursively as:

$$V_{\pi,t}(s) = R(s,\pi_t(s)) + \gamma \sum_{s' \in \mathcal{S}} T(s,\pi_t(s),s') V_{\pi,t+1}(s'), \tag{2}$$

where $\pi_t(s)$ is the action returned by the policy π at time step t given state s. Equation (2) states that the value of a given state is the immediate reward for being in that state, plus the expected discount value of the next state, assuming that the agent chooses optimally from that point forward, and V_π is the solution to the set of linear equations defined in (2). The optimal policy π^* is the one which maximizes the expected reward, and the value of the

optimal policy V_n^* starting in state s with n steps remaining can be written simply as the one which maximizes (2) over the set of actions:

$$V_n^*(s) = \arg\max_{a \in \mathcal{A}} \left[R(s,a) + \gamma \sum_{s' \in \mathcal{S}} T(s,a,s') V_{n-1}^*(s') \right]. \tag{3}$$

In this finite horizon case, the set of equations which define V^* are unique.

Many real-world applications are such that the agent is not able to observe the state directly (in the case of the problem considered here, the true identity of the target) but only through a series of observations which are a function of this unknown state; this is known as a partially observable MDP, or POMDP. The structure of a POMDP is similar to an MDP, but also encompasses:

1. a finite set of observations Ω and an observation model $O : \mathcal{S} \times \mathcal{A} \mapsto \Pi(\Omega)$ which maps states and actions to probabilities over observations. $O(s',a,o)$ specifies the probability of obtaining observation o when action a was taken, which ended up in state s',

2. a belief structure b which is a probability distribution over states $\mathcal{S}$, since the agent does not know the state in which it is. The value of $b(s)$ specifies the probability that the agent is in state s. Belief is propagated using a *state estimator* which employs Bayes' rule to update the belief when starting from state s, performing action a and obtaining observation o:

$$b^{ao} = \Pr(s'|a,o,b) = \frac{O(s',a,o) \sum_{s \in \mathcal{S}} T(s,a,s') b(s)}{\Pr(o|a,b)}, \tag{4}$$

   where

$$\Pr(o|a,b) = \sum_{s' \in \mathcal{S}} O(s',a,o) \sum_{s \in \mathcal{S}} T(s,a,s') b(s). \tag{5}$$

   is a normalization constant to ensure that $\sum_{s \in \mathcal{S}} = 1$. This state estimator function in Eqn (4) is denoted b^{ao} to indicate its dependence on the action a and observation o.

3. a transition function τ which specifies the probability of obtaining belief state b' from b via action a:

$$\tau(b,a,b') = \Pr(b'|a,b) = \sum_{o \in \Omega} \delta(b', b^{ao}) \sum_{s' \in \mathcal{S}} O(s',a,o) \sum_{s \in \mathcal{S}} T(s,a,s') b(s), \tag{6}$$

   where $\delta(x,y) = 1$ if $x = y$ and 0 otherwise.

4. a new reward function which is conditioned on the belief state:

$$\rho(b,a) = \sum_{s \in \mathcal{S}} b(s) R(s,a). \tag{7}$$

The value function for a POMDP is analogous to the MDP version from Equation (3) above, replacing appropriately to account for the partial observability:

$$V_n^*(b) = \arg\max_{a \in \mathcal{A}} \left[\rho(b,a) + \gamma \sum_{o \in \Omega} \Pr(o|a,b) V_{n-1}^*(b^{ao}) \right], \tag{8}$$

and $\Pr(o|a,b)$ is Equation (5). For this finite horizon case, it was shown [10] that $V_n^*(b)$ is a piecewise linear and convex (PWLC) function. Let ϕ represent a single linear segment of the value function and Φ the set of hyperplanes which comprise V. Since $V(b)$ is a function over belief space, can be rewritten by an $|\mathcal{S}|$-dimensional vector of ϕ coefficients:

$$V_n(b) = \max_{\phi \in \Phi} \left[\sum_{s \in \mathcal{S}} b(s) \phi(s) \right], \tag{9}$$

which, by taking the max operator, is called the max-planes representation of V. Each vector ϕ is associated to an action a, and the policy π_V for a given belief b is readily obtained using

$$\pi_V(b) = \arg\max_{a : \phi_a \in V} \phi_a \cdot b. \tag{10}$$

The concept of value functions, the max-planes representation and actions associated with ϕ vectors are illustrated in Figure 1.

Equation (9) is formulated for the finite horizon case. In many instances, however, it is not desirable, or possible, to specify the horizon k in advance. In the case of $k = \infty$ one obtains an infinite horizon POMDP. In this instance, although V remains convex, it may have an infinite number of facets. However,

$$\lim_{n \to \infty} \|V_n^* - V^*\| = 0, \tag{11}$$

does hold and it is possible to approximate the infinite horizon value function arbitrarily close by using a sufficiently long horizon.

Computing the value function and policy for a POMDP is a difficult task owing to the fact that the belief space is an infinite continuous space over $\mathcal{S}$. Techniques such as value iteration [9] are intractable for some problems. In order to give good solutions, methods have been developed which maintain upper and lower bounds on V^*, using point-based updates instead of the full Bellman update [11], thus improving only a small subset of the

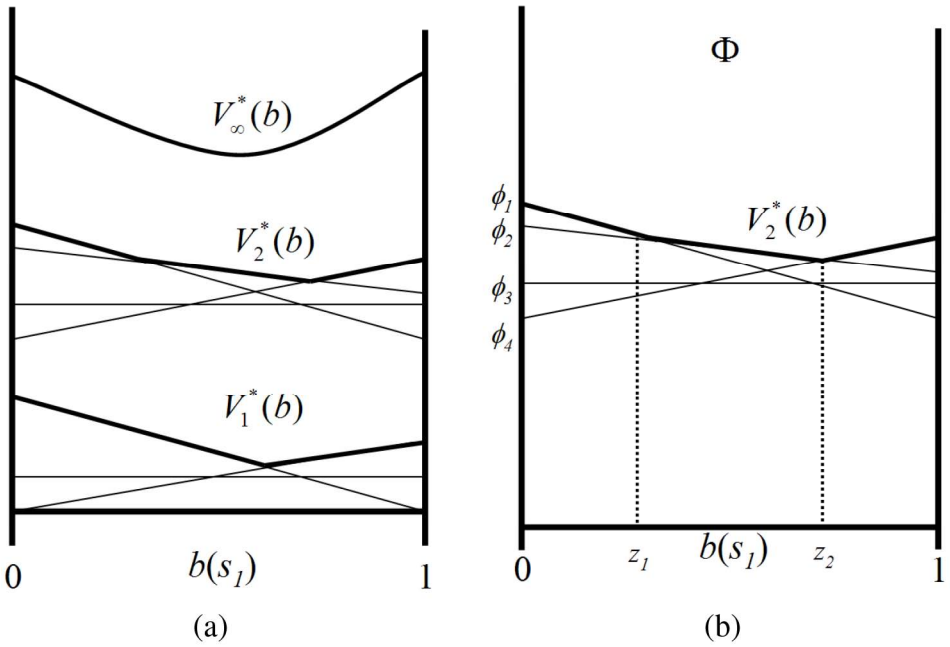


Figure 1: *A graphical representation of a two-state POMDP to help provide some geometrical intuition of the optimal value function. Since* $\sum b(s) = 1$*, the belief space is depicted with a single dimension, and the value function is a curve. Panel (a) shows the optimal value function* $V^*(b)$ *for various horizons* k*. The thin lines show the* ϕ*-vectors which represent a particular policy of horizon length* k *and the upper surface, the max-planes representation of* V*, is shown with the thicker line. The number of* ϕ *used to represent* V *usually grows with larger values of* k*; however it remains piecewise-linear and convex. For* V_∞^**, the surface remains convex but may require an infinite number of* ϕ *vectors and is no longer piecewise linear. Panel (b) shows how an action is derived from the* ϕ *vectors for the* $k = 2$ *value function. The set* Φ *contains four* ϕ *vectors, but only three of them are used in the max-planes representation of* $V_2^*(b)$*. For belief values* $[0 \ldots z_1]$*, the action given by the policy associated with* ϕ_1 *is optimal; from* $[z_1 \ldots z_2]$ *then the action of* ϕ_2 *is optimal and a belief in* $[z_2 \ldots 1]$ *will cause the agent to execute the action associated with* ϕ_4*.*

state space at a time but without the expense, and more importantly incorporate information in the form of heuristics to guide the search to where to apply the point-based updates, and ignoring irrelevant areas of the search space. The method used in this study, developed by Smith, is called Focused Real-Time Dynamic Programming (FRTDP) [12]. Details are given in [13].

3 A POMDP for target classification

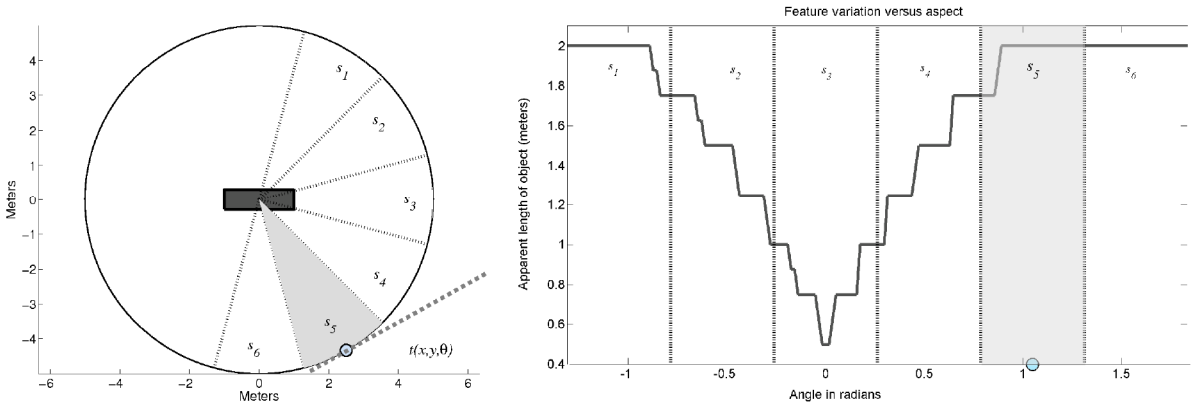
With the foundation of a POMDP having been laid out previously, a target classification problem which contains the necessary components will be defined that will allow the evaluation of the value function and the corresponding policy. The policy can then be applied to determine the behaviour of the vehicle performing the classification of the object, such as generating waypoints to obtain an additional aspect. The decision cycle is summarized as follows:

1. With the current belief state b, perform the action given by the policy $a = \pi^*(b)$
2. Obtain an observation o
3. Update the belief state with the state estimator in Equation (4).

3.1 Model

Figure 1 shows the geometry involved in the sidescan problem. The agent (here, the autonomous vehicle) detects the presence of a target with its sidescan sonar. This results in a number of measurements which which give some indication as to the nature of the target. The vehicle is now faced with a decision as to whether it should classify the target with the information it has or to obtain another aspect on the target to reduce the uncertainty. The set of states $\mathcal{S}$ contains each possible aspect for each target and non-target class, resulting in the cardinality $|\mathcal{S}| = n_t \cdot n_a$, where n_t is the number of objects and n_a is the number of possible aspects. We denote s_{ta} as the state associated with target t and aspect a, and $s_{\bar{t}a}$ as the state associated with the non-target class $\bar{t}$. The states are shown in Figure 1 (a). The symmetry of the observations used for this simulated example means that aspects at $\pm 180^o$ are redundant and thus removed to reduce the size of the problem; in general, this may not be the case and depends on the observation model used. This simulated problem contains two target types: a cylindrical object (shown in Figure 1) and a smaller wedge-shaped target. For the number of aspects, the range over π is discretized over 6 equally sized angular increments. The non-target class is modeled as an ellipse with random major and minor axes which are uniformly distributed over the range of the size of the targets.

When a vehicle equipped with a sidescan sonar detects a target, it obtains a measurement in the form of a feature vector $\mathbf{x}$ which is computed from the sonar imagery. The observation set Ω must contain features which vary appropriately with aspect. In this simulation, the feature used is the apparent length of the target as seen from the sensor's viewpoint. The probability of each observation given a target and aspect $P(o|s_{ta})$ is modeled using the histograms of the length of the modeled targets. The variation of o versus aspect (and state) for the cylindrical object is shown in Figure 1(b). The curve given by this feature is not smooth since the sensor has a finite resolution, causing discontinuities at resolution boundaries. For the non-target class, $P(o|s_{\bar{t}*})$ is modeled using the uniform distribution



(a) (b)

Figure 2: *A sketch of the problem setup. Panel (a) shows a cylindrical target which can be sensed at a number of different aspects. The circle shows the 5 meter range ring to give an indication of scale. A vehicle at a given aspect angle (shown as a dot) traveling on the path shown with the dotted line, detects a target at a 5 meter range with its sidescan sonar in the direction perpendicular to the track. The aspects are discretized in the number of states, here* $\mathcal{S} = [s_{c1} \ldots s_{c6}]$*, where* c *indicates that the target is a cylinder. The observation used in this test example is the apparent length of the object as measured from the sonar imagery. This is shown in panel (b) as it varies with aspect. The lines delimiting the states are also shown. In this example, the vehicle would obtain an observation* $o = 2$ *meters and would be used to update the belief function. Due to the symmetry of this feature, only the* $[0 \ldots \pi]$ *angles are shown.*

over o. The FRTDP technique requires a discrete number of observations. To reduce the size of the problem and generate a finite and discrete number of observations, the range of measurements is quantized using a simple code-book generated using a k-means clustering method (e.g [14]). For this example, $|\Omega| = 4$.

The vehicle has a number of options as to what action to perform given a belief state $b(s)$. The action set $\mathcal{A}$ contains the instructions which direct the vehicle to obtain the aspects to gain information, and actions which are declarations that cause the vehicle to decide as to the nature of the object. Since there are 6 possible aspects at which a target may be viewed, $\mathcal{A}$ contains 5 actions, denoted using degrees for ease of comprehension, which move the vehicle to one of the other sectors,

$$\{\text{Move\_Plus\_30}, \text{Move\_Plus\_60}, \text{Move\_Plus\_90}, \text{Move\_Plus\_120}, \text{Move\_Plus\_150}\}.$$

In addition, there are three declaration actions

$$\{\text{Declare\_Cylinder}, \text{Declare\_Wedge}, \text{Declare\_NonTarget}\},$$

resulting in $|\mathcal{A}| = 8$. The transition matrix in this case is simple, since in general vehicle navigational accuracy is precise enough as to make the transitions probabilities deterministic: for instance in the example shown in Figure 2, if the true state is s_{c5}, the action Move_Plus_30 put the system in state s_{c6} with probability 1.0. Because of the symmetry of the feature used, there is a wrap-around effect which causes, for example the action Move_Plus_30 in state s_{*6} to transition into state s_{*1}, where the wildcard $*$ is to show that this is applicable to every object.

It is via the reward function $R(s,a)$ that the agent's behaviour can be most influenced, and there is a great deal of flexibility in its specification. In general, some compromise between classification accuracy and the costs associated with missed detections and false alarms is needed. For instance, if one makes the cost (a negative reward) of missing a target too high, the agent will always declare target, in order to avoid the possibility of incurring this large penalty. If one makes the cost associated with obtaining a new aspect too low, then the agent may attempt to acquire many or all aspects of the object in order to reduce its uncertainty to such a small number as to negate the benefits of the technique. Keeping this in mind, it was decided that, for the numerical simulation of this section, maximum classification accuracy was desired in order to test the developed concepts: thus, $R(s,a)$ was equal for all declaratory actions in which the state was correct, and equally negative for incorrect classifications (in this case ± 10). The cost of obtaining an aspect was $1/10^{\text{th}}$ that of the reward value (here -1).

3.2 Numerical simulation

The Focused Real-Time Dynamic Program, as implemented in the ZMDP software suite [12] was used to compute the infinite horizon value function V^* for the POMDP defined earlier. This is then used to obtain the corresponding policy π^* defined by the value function. A policy was found with the maximum regret, defined as the difference between the expected reward of the optimal policy and the current policy [15], being less than 0.0009058 for the rewards state above (note that the computations were stopped once the regret reached a level below 0.001). A total of 145 $\phi-$ vectors were used to create the max-planes representation of V^*. With the computed policy π^*, a numerical simulation was carried out whereby, at each iteration one of the three object classes was randomly generated at a random aspect. The vehicle obtains the first measurement "for free" since the sonar image upon which detections are made is the same image used by the classification method; thus, the agent begins the process with an already updated belief structure $b(s)$, based on the observation received using Equation (4). The confusion matrix for 1000 iterations is shown in Table 1. Shown in Table 2 is the result of using only two perpendicular aspects, a technique called cross-hatching which is often used in practice. Figure 3.2 shows the vehicle path for an example classification of one of the non-target objects using the policy derived above. The order in which the aspects were obtained are numbered. A small "lead-in" and "lead-out" factor was applied to simulate a vehicle which would line

		Predicted		
		c	w	n
True	c	337	0	1
	w	0	303	0
	n	19	101	239

Table 1: *Confusion matrix for cylinder (c), wedge (w) and non-target (n) classes over 1000 iterations using the POMDP defined above.*

		Predicted		
		c	w	n
True	c	307	0	31
	w	0	303	0
	n	11	169	179

Table 2: *Confusion matrix for cylinder (c), wedge (w) and non-target (n) classes over 1000 iterations using only two perpendicular aspects.*

itself up onto a straight track in order to obtain a satisfactory image. Underwater vehicles are in general non-holonomic and as a consequence will require a non-zero turning radius to follow a path between two points, especially if they are close together. Here a turning radius of 2 meters was used and a Dubins turn [16], which defines the shortest path between two points and headings that an AUV is able to follow [1] was computed. Since the vehicle has a choice of sensing the target from the port or starboard side, a simple greedy optimization chooses the path which minimizes the total distance traveled to obtain the next aspect.

3.3 Discussion

Tables I and II show that the policy obtained using the value function of the defined POMDP is more accurate than using only the perpendicular views of the target, especially while classifying the problematic non-target class. The policy ”requested” on average 2.24 views (the breakdown was 1.77 when the class was a cylinder, 2.22 when it was a wedge and 2.69 when it was a non-target). However, the second view was overwhelmingly the perpendicular view. This is not surprising, since one would expect that using this feature would lead to the situation where the two perpendicular views give independent samples on the target. The power of the method lies in the ability not only to resolve this, but also to optimize the choice of the follow-on aspects. With roughly one additional look required for each four contacts detected, significant gains can be obtained in classification performance.

[1] Thanks to M. Couillard, DRDC-CORA for this implementation

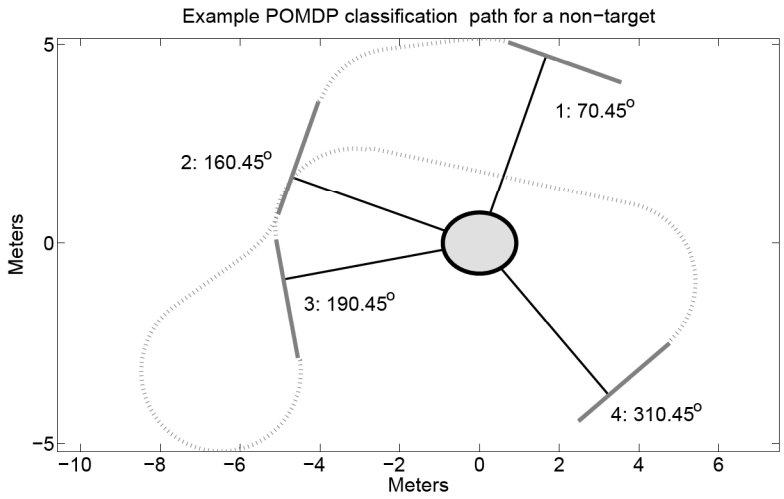


Figure 3: *The path the simulated vehicle would use during the execution of the policy. The vehicle path is shown as dotted lines for transit paths, and solid lines for sensing paths (with lead-in and lead-out distances). The aspect to the target is shown and the order (with the aspect in degrees) is also shown. The classification of this non-target object required 4 aspects.*

A nice consequence of the way the problem has been posed is the definition of a generic non-target class which gives the agent the option of classifying an object as *not* one of known target classes. In this case, the uniform distribution of the features for this category imposes the lowest amount of information about the class. Of course, if one had greater knowledge about the non-target class then it should be integrated into the the observation model O; without such knowledge, the uniform distribution is the most sensible choice. In this simulation, the non-target class did indeed follow the uniform distribution for the possible observations. If this was not the case, some performance would be sacrificed (Proportional to the amount of mutual information between the two distributions (e.g. assuming $P_1(x)$ when it in fact the true distribution is $P_2(x)$) .

4 Conclusion

This memo has introduced a partially observable Markov decision process for multi-aspect target classification using AUVs equipped with imaging sidescan sonar. The parameters required for the problem, namely the actions available to the vehicle, the states for the targets and the state transition probabilities were specified in a generic way. The observation set and probabilities were created using a simple feature, the apparent object length, with a view of proving the concept but which is unlikely to work as well on real data. The resulting POMDP provides a predictable method computing the next best aspect angle to obtain and the resulting waypoints can be used to provide a reactive behaviour capability in response to the incoming sensor data to the vehicle control system. The flexibility afforded by the parameters gives the system designer the ability to specify realistic values for cost and rewards which affect overall mission progress. In addition, not requiring the explicit definition of the non-target class provides a more realistic method of specifying the clutter class, where one is likely to generally know the properties of the target class but not the non-target class. This is especially true for geometric features obtained with high-frequency imaging sonar.

5 Future Work

Several options are available for future development, the most important of which is to apply the technique to real sonar data. The sonar data, however, must be of sufficient quality as to be able to properly quantify the variation with aspect of the targets under consideration. A set of very high resolution synthetic aperture sonar (SAS) data is being processed to this end. While much of the POMDP framework developed here will remain the same, a more robust model-based feature is being developed as the observation model. Indeed, the framework used here defines discrete observations and actions, whereas in reality these values can be continuous, especially the computed features. POMDPs defined on a continuous space, such as PERSEUS [17] could be investigated, however it is not clear whether the potential gains in performance will justify the increased complexity. Another avenue for investigation is rather than explicitly define some of the components of the POMDP such as the reward or the observation model, these could be learned as the AUV interacts with its environment through reinforcement learning [18]. The drawback in this case, however, is that the expense and risk associated with employing this technique during operations or experiments at sea would likely mean that a great deal of the learning would be performed in simulation, thus negating the more desirable aspects of this method.

References

[1] Reed, S., Petillot, Y., and Bell, J. (2004), Model-Based Approach to the Detection and Classification of Mines in Sidescan Sonar, *Applied Optics*, 43(2), 237–246.

[2] Couillard, M., Fawcett, J.A., Davison, M., and Myers, V. (2007), Optimizing Time-Limited Multi-aspect Classification, In *Proceedings of the Institute of Acoustics*, Vol. 29, pp. 89–96.

[3] Zerr, B., Fawcett, J., and Hopkin, D. (2009), Adaptive Algorithm for Sea Mine Classification, In *Underwater Acoustic Measurements (UAM) Conference Proceedings*.

[4] Fawcett, J.A., Crawford, A., Hopkin, D., Myers, V., Couillard, M., and Zerr, B. (2008), Multi-aspect computer-aided classification of the Citadel trial sidescan sonar images, Technical Report Defence Research and Development Canada.

[5] Williams, D. and Groen, J. (2009), Multi-view target classification in synthetic aperture sonar imagery, Technical Report Underwater Acoustic Measurements (UAM) Conference Proceedings.

[6] Cassandra, A. (1998), A survey of POMDP applications, *AAAI fall symposium*.

[7] He, L., Ji, S., and Carin, L. (2006), Application of Partially Observable Markov Decision Processes to Robot Navigation in a Minefield, In *ICAPS*.

[8] Kaelbling, L. P., Littman, M. L., and Cassandra, A. R. (1998), Planning and acting in partially observable stochastic domains, *Artificial Intelligence*.

[9] Russell, S. and Norvig, P. (2002), Artificial Intelligence: A Modern Approach, Prentice-Hall.

[10] Smallwood, R. D. and Sondik, E. J. (1973), The Optimal Control of Partially Observable Markov Processes over a Finite Horizon, *Operations Research*, pp. 1071–1088.

[11] Bellman, R.E. (1957), Dynamic Programming, Princeton University Press.

[12] Smith, T. (2009), ZMDP Software for POMDP and MDP Planning. http://www.cs.cmu.edu/ trey/zmdp/.

[13] Smith, T. (2007), Probabilistic Planning for Robotic Exploration, Ph.D. thesis, Carnegie Mellon Institute.

[14] Gersho, A. and Gray, R. M. (1991), Vector Quantization and Signal Compression, Kluwer.

[15] Smith, T. and Simmons, R. (2004), Heuristic Search Value Iteration for POMDPs, In *Proc. Int. Conf. on Uncertainty in Artificial Intelligence (UAI).*

[16] Dubins, L.E. (1957), On curves of minimal length with a constraint on average curvature, and with prescribed initial and terminal positions and tangents, *American Journal of Mathematics*, 79, 497–516.

[17] Porta, J. M., Vlassis, N., Spaan, M.T.J., and Poupart, P. (2006), Point-Based Value Iteration for Continuous POMDPs, *Journal of Machine Learning Research*, 7, 2329–2367.

[18] Sutton, R. S. and Barto, A. G. (1998), Reinforcement Learning: An Introduction, MIT Press.

This page intentionally left blank.

Distribution list

DRDC Atlantic TM 2009-154

Internal distribution

1 Author

1 Mae Seto

1 John Fawcett

1 Warren Connors

3 Library

Total internal copies: 7

External distribution

1 DRDKIM

1 Library and Archives Canada

Total external copies: 2

Total copies: 9

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.)

   Defence R&D Canada – Atlantic
   P.O. Box 1012, Dartmouth, Nova Scotia, Canada
   B2Y 3Z7

2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.)

   UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

   Adaptive target classification with imaging sonar as a partially observable Markov decision process

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used.)

   Myers, V.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total containing information. Include Annexes, Appendices, etc.)	6b. NO. OF REFS (Total cited in document.)
January 2010	30	18

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

   Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

   Defence R&D Canada – Atlantic
   P.O. Box 1012, Dartmouth, Nova Scotia, Canada B2Y 3Z7

9a. PROJECT NO. (The applicable research and development project number under which the document was written. Please specify whether project or grant.)	9b. GRANT OR CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
11cf	

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.)	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)
DRDC Atlantic TM 2009-154	

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

   (X) Unlimited distribution
   ( ) Defence departments and defence contractors; further distribution only as approved
   ( ) Defence departments and Canadian defence contractors; further distribution only as approved
   ( ) Government departments and agencies; further distribution only as approved
   ( ) Defence departments; further distribution only as approved
   ( ) Other (please specify):

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11)) is possible, a wider announcement audience may be selected.)

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

Discriminating between different types of objects on the seabed using high-resolution imaging sonar is challenging, in part due to the inference of a 3-dimensional shape using one or more 2-dimensional projections. Often, more than one aspect of a detected object is required in order to correctly classify it as one of a number of targets of interest such as an influence mine, or as a benign non-target. When the sensor is equipped by an autonomous underwater vehicle (AUV), the acquired aspects are usually determined by a pre-planned mission rather than in response to the sensor data (i.e. the process is purely deliberative rather than reactive). In this paper, the multi-aspect target classification problem is modeled as a partially observable Markov decision process (POMDP). The solution to a POMDP is called a policy which provides a means for an AUV's control system to determine a course of action in response to the incoming sensor data, such as classifying an object as a target or not, or obtaining data from a specific aspect to reduce the uncertainty. The components of a POMDP for multi-aspect object classification with an AUV equipped with a sidescan are formulated. A numerical simulation is undertaken to assess the performance of the resulting policy and the vehicle path which is reacting to simulated sensor data is shown. The simulation was performed using two targets (a cylindrical object and a small, wedge shaped object) and a non-target class modeled as an elliptical object with uniformly distributed major and minor axes. The resulting policy was executed for many iterations and compared to one which obtains two perpendicular aspects. The classification accuracy of the POMDP model, as measured using a confusion matrix, was markedly superior to the cross-hatching strategy, while obtaining an average of 2.22 aspects (versus 2 for cross-hatching).

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus. e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Autonomy
AUV
Automatic Target Recognition
POMDP
MCM

This page intentionally left blank.

Defence R&D Canada

Canada's leader in defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

www.drdc-rddc.gc.ca